

Commentary—Goodbye M(C)ID! Hello MID, Where Do You Come From?

Holger J. Schünemann and Gordon H. Guyatt

Health-related quality of life (HRQL) has overcome many barriers that limited its acceptance as an important outcome in health care. One of the remaining barriers relates to how one should interpret HRQL scores when they change over time within patients, or differ between patients. Indeed, because there is no gold standard methodology, interpreting HRQL scores is a challenging task.

While daunting, addressing the challenge of interpretability is crucial to moving the field forward. The choice of what constitutes an important difference in a HRQL score will influence judgments about the success of a health care intervention, the required sample size of clinical studies, and the design of these studies. The issue is relevant to clinicians, payers, funding agencies, and regulatory agencies, and most relevant to patients for whose health care these groups claim responsibility.

Several approaches to assessing interpretability exist. Anchor-based approaches rely on examining the relation between scores on a HRQL instrument that is under investigation and an anchor, an independent measure of HRQL that clinicians can easily interpret (Guyatt et al. 2002). Other approaches for evaluation of interpretability of HRQL scores include distribution-based or statistical methods and reliance on experts (panel-based methods) (Lassere et al. 2001). Wyrwich and colleagues utilized the last of these alternatives to determine interpretability of the SF-36 by elegantly combining outcomes research with qualitative research methods (Wyrwich et al. 2003, 2005). This approach focuses on how clinician researchers view patients in relation to their HRQL scores and changes in HRQL scores. The article by Wyrwich's et al. is ingenious for several reasons.

First, despite the wealth of literature evaluating the SF-36 as an outcome measure, evidence for the interpretability of the instrument is surprisingly limited. Second, Wyrwich and colleagues focus on three different clinician groups and estimates whether interpretability differs across these groups. Third, the judgments the clinician researchers made were based on detailed patient scenarios.

In this commentary we will provide two arguments that readers should consider in the context of the work by Wyrwich and colleagues. First, we argue

that research on the interpretability of HRQL instruments should focus primarily on the patient's view. Early work by our group pioneered the methodology of assessing interpretability of HRQL instruments (Jaeschke, Singer, and Guyatt 1989). In that work, we described what became widely known as the minimal clinical important difference (MCID). Because this terminology focuses attention on the clinical arena rather than patients' experience in their day-to-day lives, we subsequently removed the focus on "clinical" interpretations, and the "C" from MCID to focus on the minimal important difference (MID) (Juniper et al. 1994; Schünemann et al. 2005). HRQL is a patient important outcome because it is the patients who experience their HRQL, and only they are in a position to ultimately judge whether a difference is important (Guyatt et al. 2004).

We now define the MID as the smallest difference in score in the outcome of interest that informed patients or informed proxies perceive as important, either beneficial or harmful, and which would lead the patient or clinician to consider a change in the management (Schünemann et al. 2005). The revised description of the MID precludes making MID estimates for outcomes that are remote from those important, in themselves, to patients, such as spirometry or laboratory exercise capacity. Further, the definition suggests that only if one had reason to question the reliability or accuracy of data from patients would one rely on proxies to provide estimates of the MID.

Investigators or clinicians may differ in the perspective or the methodology they adopt to determine the MID. Thus, readers, when they interpret the results of research on the MID, must attend to who rated the importance of an HRQL change and the specific instructions provided for making the assessment. If one accepts that HRQL measurement must be fundamentally patient-centered, the first choice for establishing the MID should be a patient-based approach. Relative to patients, clinicians may overemphasize treatment effects (Puhan et al. 2004) and agreement between patients and proxies in rating of HRQL is not perfect (Sneeuw, Sprangers, and Aaronson 2002; Ubel, Loewenstein, and Jepson 2003; von Essen 2004). Again, if one accepts that patients are at the center in HRQL measurement, then investigators, when

Address correspondence to Holger J. Schünemann, M.D., Ph.D, M.Sc., CLARITY GROUP, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada. Dr. Schünemann and Gordon H. Guyatt M.D., MSc. are with the CLARITY Group, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada. Dr. Schünemann is also with the Department of Medicine, University of Buffalo, New York, USA and INFORMA, National Cancer Institute, Rome, Italy. Dr. Guyatt is also with the Department of Medicine, McMaster University, Hamilton, Ontario, Canada.

they do use proxies, should instruct those proxies to focus on what they believe patients consider important.

Wyrwich and colleagues took a different approach. “Panelists were not provided with any specific definition of a CID (clinically important difference), but left to determine their own meaning for this term” (p. 580). Thus, the investigators did not specify to the participants that they were to estimate the difference that patients consider important. Further, they used the term “CID,” further permitting ambiguity about the group to whom the difference is important.

Our second argument concerns the results of the study. If readers dismiss our appeal to focus on the patient, they should examine the result of the study. While the results, including the difference in the MID between heart failure clinician researchers and those working in asthma or chronic obstructive pulmonary disease (COPD) provide intriguing insights into clinicians’ perspectives, the extent to which they enlighten us concerning differences in SF-36 scores that patients consider important remains questionable. At least two issues are worth considering. First, readers should ask whether the differences in the MID for the three evaluated diseases are true. Methodological issues, such as sampling bias in terms of the physician researchers selected for this exercise, or randomness around the consensus estimates could explain these MID differences. There were no statistical approaches—and perhaps there are none for this study design—to compare the magnitude of the MID differences across diseases obtained from the physician researchers. Although Wyrwich and colleagues performed a laborious study, readers should ask how large would the variation around the obtained MID estimates be if they had conducted multiple focus groups for each disease category. Could chance explain the observed differences in the MIDs from the three focus groups? Readers should not ignore this possible explanation.

Second, if the MID on the SF-36 was truly greater for heart failure patients compared with patients suffering from asthma or COPD, it might also have implications for the comparability of SF-36 scores. The SF-36 is a generic HRQL instrument. To be applicable to different patient populations, similar scores across different patient population should signify similar levels or impairment of HRQL. For instance, a mean score of 50 on the 0–100 scale of the physical functioning domain should indicate a similar level of physical function for different patient populations such as patients with asthma, COPD, and heart failure. If interventions caused an improvement of HRQL to 60, the result of the study by Wyrwich and colleagues would suggest that the change is important for the former two groups but not the latter. Would the

post-intervention scores of 60 still be comparable across these populations? It is conceivable that differences between patient groups in what constitutes an important change could impair the comparability of generic instruments. Thus, the implications of an MID for the SF-36 that varies by population could be far reaching. Such results could question the SF-36 as a generic instrument that allows a straightforward comparison across diseases.

In summary, the MID provides an important strategy to make interpretable the results of HRQL studies. To be maximally informative, representative samples of informed patients or their proxies should provide estimates of the MID.

REFERENCES

- Guyatt, G. H., D. Osoba, A. W. Wu, K. W. Wyrwich, G. R. Norman, and Clinical Significance Consensus Meeting. 2002. "Methods to Explain the Clinical Significance of Health Status Measures." *Mayo Clinic Proceedings* 77 (4): 371–83.
- Guyatt, G., V. Montori, P. J. Devereaux, H. Schünemann, and M. Bhandari. 2004. "Patients at the Centre: In Our Practice, and in Our Use of Language." *ACP Journal Club* 140: A-11.
- Jaeschke, R., J. Singer, and G. H. Guyatt. 1989. "Measurement of Health Status. Ascertaining the Minimal Clinically Important Difference." *Controlled Clinical Trials* 10 (4): 407–15.
- Juniper, E. F., G. H. Guyatt, A. Willan, and L. E. Griffith. 1994. "Determining a Minimal Important Change in a Disease-Specific Quality of Life Questionnaire." *Journal of Clinical Epidemiology* 47 (1): 81–7.
- Lassere, M. N., D. van der Heijde, and K. R. Johnson. 2001. "Foundations of the Minimal Clinically Important Difference for Imaging." *Journal of Rheumatology* 28: 890–1.
- Puhan, M. A., M. Behnke, P. J. Devereaux, V. M. Montori, O. Braendli, M. Frey, and H. J. Schünemann. 2004. "Measurement of Agreement on Health-Related Quality of Life Changes in Response to Respiratory Rehabilitation by Patients and Physicians—A Prospective Study." *Respiratory Medicine* 98 (12): 1195–202.
- Schünemann, H., M. Puhan, R. Goldstein, R. Jaeschke, and G. Guyatt. 2005. "Measurement Properties and Interpretability of the Chronic Respiratory Disease Questionnaire (CRQ)." *Journal of COPD* (in press).
- Sneeuw, K. C. A., M. A. G. Sprangers, and N. K. Aaronson. 2002. "The Role of Health Care Providers and Significant Others in Evaluating the Quality of Life of Patients with Chronic Disease." *Journal of Clinical Epidemiology* 55 (11): 1130–43.
- Ubel, P., G. Loewenstein, and C. Jepson. 2003. "Whose Quality of Life? A Commentary Exploring Discrepancies between Health State Evaluations of Patients and the General Public." *Quality Life Research* 12: 609–10.

- von Essen, L. 2004. "Proxy Ratings of Patient Quality of Life—Factors Related to Patient-Proxy Agreement." *Acta Oncology* 43 (3): 229–34.
- Wyrwich, K. W., S. D. Fihn, W. M. Tierney, K. Kroenke, A. N. Babu, and F. D. Wolinsky. 2003. "Clinically Important Changes in Health-Related Quality of Life for Patients with Chronic Obstructive Pulmonary Disease. An Expert Consensus Panel Report." *Journal of General Internal Medicine* 18 (3): 196–202.
- Wyrwich, K. W., W. M. Tierney, A. N. Babu, K. Kroenke, and F. D. Wolinsky. 2005. "A Comparison of Clinically Important Differences in Health-Related Quality of Life for Patients with Chronic Lung Disease, Asthma or Heart Disease." *Health Services Research* 40 (2): 577–92.